基于深度循环神经网络的跨领域文本情感分析*

余传明
中南财经政法大学信息与安全工程学院 武汉 430073

摘要：[**目的/意义**]通过在标注资源丰富的源领域(Source Domain)中学习，并将目标领域(Target Domain)的文档投影到与源领域相同的特征空间(Feature Space)中去，从而解决目标领域因标注数据量较小而难以获得好的分类模型的问题。[**方法/过程**]选择亚马逊在书籍、DVD 和音乐类目下的中文评论作为实验数据，以跨领域情感分析作为研究任务，提出一种跨领域深度循环神经网络(Cross Domain Deep Recurrent Neural Network，CD-DRNN)模型，实现不同领域环境下的知识迁移。CD-DRNN 模型在跨领域环境下的平均分类准确度达到了81.70%，优于传统的栈式长短时记忆网络(Stacked Long Short Term Memory，Stacked-LSTM)模型(79.90%)、双向长短时记忆网络模型(Bidirectional Long Short Term Memory，Bi-LSTM)模型(80.50%)、卷积神经网络长短时记忆网络串联(Convolution Neural Network with Long Short Term Memory，CNN-LSTM)(74.70%)模型以及卷积神经网络长短时记忆网络并联(Merged Convolution Neural Network with Long Short Term Memory，Merged-CNN-LSTM)模型(80.90%)。[**结果/结论**]源领域和目标领域的知识迁移能够有效解决监督学习在小数据集上难以获得好的分类效果的问题，通过 CD-DRNN 模型能够从无标注数据中有效地筛选特征，从而大大降低目标领域数据标注相关的工作量。

关键词：跨领域 迁移学习 深度循环神经网络 情感分析
分类号：TP391
DOI：10.13266/j.issn.0252-3116.2018.11.003

随着社交网络(Social Networks)的发展与深入，越来越多的人倾向于通过社交网络表达和分享自己对某些产品的主观意见，同时，在购买产品前获取和分析相应的网络评论作为重要参考意见。从文学作品(例如书评、影评和音乐评论)到电子产品(例如手机、数码相机和家电产品)，从吃喝玩乐(例如餐饮、景点评论和旅游评论)到衣食住行(例如服装评论、地产评论和汽车评论)，评论所涉及的领域往往分布广泛并且比例失衡。在自动化的信息分析过程中，如果对所有领域进行单独建模，则需要耗费大量的时间和人力。鉴于此，研究者希望在一些相似或者相近的领域之间建立迁移学习(Transfer Learning)[1]或者领域适配(Domain Adaptation)[2]模型。例如，通过标注书评领域的数据，针对所学习到的情感分类模型或者主题分布模型，将其迁移或者适配到影评领域，从而节省大量影评领域的标注时间。

在迁移学习(领域适配)的实际应用中，往往存在较为严重的概念漂移(Concept Drift)[3]和特征漂移(Feature Drift)[4]。例如，"酸甜苦辣"这类表征味觉的特征词在美食评论领域的分布频率远高于其在家电领域的分布频率；特征分布在不同领域所呈现的巨大差异导致情感分类模型在跨领域应用中会有所下降。另外，在不同的领域，同样的特征可能揭示不同的情感极性。例如，在餐饮评论领域，"时间长"倾向于表达等待上菜时间过长这一负面情感；在电子产品评论领域，"时间长"则更倾向于表达电池待机时间长这一正向情感。这导致餐饮领域的情感模型在应用到电子产品评论领域时，情感分类效果急剧下降。

为了解决领域跨度环境下的概念漂移(特征漂移)问题，本文选择亚马逊在书籍、DVD 和音乐类目下的中文评论作为实验数据，情感分析作为研究任务，提出一种面向跨领域的深度循环神经网络模型(Cross

* 本文系国家自然科学基金面上项目"大数据环境下基于领域知识获取与对齐的观点检索研究"(项目编号：71373286)和国家自然科学基金青年项目"突发公共卫生事件社交媒体信息主题演化与影响力建模"(项目编号：71603189)研究成果之一。
作者简介：余传明(ORCID：0000-0001-7099-0853)，副教授，E-mail：yucm@zuel.edu.cn。
收稿日期：2017-09-06 修回日期：2018-03-04 本文起止页码：23-34 本文责任编辑：王善军

Domain Deep Recurrent Neural Network, CD-DRNN),以实现不同领域环境下的知识迁移。

1 相关研究

所谓跨领域文本情感分析,是指特征分布在不同领域存在差异的情况下,利用源领域和目标领域之间的联系,将源领域的相关知识迁移到目标领域中,通过机器学习算法实现对目标领域的文本情感分析。从研究视角来看,其内容涉及不同领域之间的知识迁移[5]、机器学习[6]和情感计算[7]等研究问题,以下分节展开论述。

1.1 知识迁移视角的跨领域文本情感分析

从知识迁移的视角来看,跨领域文本情感分析主要解决领域跨度环境下的特征分布差异,以及由此产生的概念漂移(特征漂移)问题,其方法包括特征迁移(特征适配)、实例迁移(实例适配)和模型迁移(模型适配)等。

特征迁移着力于解决源领域与目标领域特征空间不一致的问题,通过特征变换将目标领域数据的特征向量映射到与源领域相同的向量空间中,从而实现迁移学习。例如,J. Blistzer 等人提出结构对等学习(Structural Correspondence Learning, SCL)模型[8],首先在多领域之间筛选出一些具有共性的特征词(即枢纽词汇,又称为 Pivot Words),然后量化其与正负向情感表达之间的关系,并利用传统的机器学习模型对量化关系进行建模。Pan 等人提出谱特征对齐(Spectral Feature Alignment, SFA)方法[9],将文本特征区分为领域相关特征和领域无关特征;以领域无关特征作为主线,将领域相关的特征聚类到与其距离最小的领域无关特征之上。上述方法依赖于枢纽词汇(或领域无关词汇)在不同领域的分布情况,若其分布过于稀疏,则很难获得较好的效果。

实例迁移是指从源领域中尽可能筛选出与目标领域相似的标注数据,从而增大目标领域的训练数据,并将其应用到目标领域的情感分析任务之中。例如,W. Dai 等人提出了 TrAdaboost 方法,在给定目标领域少量标注文本的情况下,通过利用大量源领域的标注数据,构建了高质量的文本分类模型[10]。马风闸等人提出了双重策略选择策略,根据源领域的样本权重进行首次选择,再结合分类置信度对样本进行二次选择。其实验结果表明,该模型有助于提高整体分类精度[11]。实例迁移方法依赖于源领域标注样本的数量,若其数量较小,则很难适用。

模型迁移是指在源领域和目标领域共享模型或者参数,从而实现针对目标领域的文本情感分析。例如,X. Glorot 等人利用一种深度学习模型,即降噪自编码器(Stacked Denoising Autoencoder, SDAe)对多领域的无标注数据进行预训练,利用预训练模型结合源领域的有标注文档来训练情感分类模型,该模型在 22 个领域的分类效果优于 SCL 模型与 SFA 模型[12]。尽管如此,SDAe 算法存在速度较慢、且较多地依赖于参数初始化等问题。有研究者对其改进,提出了 MarginalizedSDAe 算法[13],使其在运行速度上得到了提高。由于模型迁移方法减少了对枢纽词汇(或领域无关词汇)和源领域标注数据的依赖,在跨领域文本情感分析的效果上,结合深度学习的模型迁移总体效率上优于传统的特征迁移和样本迁移。

1.2 机器学习视角的跨领域文本情感分析

从机器学习的视角来看,跨领域文本情感分析主要包括基于图模型的方法、传统统计学习方法和基于深度学习的方法等。基于图模型的方法通常将源领域和目标领域的词汇抽象为节点,将词汇之间的关系(例如共现关系)抽象为边,通过一定算法计算节点之间的相似度,借此实现对目标领域的情感分析。SimRank 算法[14]是最为典型的图模型方法,其基本思想是:节点与其本身的相似度最高;如果两个节点具有更多相同或者相似的邻居节点,则其具有更高的相似性。魏现辉等人对 SimRank 进行改进,提出了加权 SimRank 算法[15],并将其应用于跨领域情感分析,实验结果验证了加权模型的有效性。

传统统计学习方法通常在给定源领域和目标领域文本的情况下,估计目标领域文本属于某个情感类别的条件概率。例如,张志武[16]将逻辑回归模型(Logistic Regression, LR)用于跨领域情感分析,实验结果表明了该模型的有效性。黄瑞阳等人[17]比较了朴素贝叶斯方法(Naïve Bayes, NB)、支持向量分类方法(Support Vector Classification, SVC)和期望最大值方法(Expectation Maximization, EM)在跨领域情感分析中的应用。其实验结果表明,EM 方法效果略优于 NB 方法和 SVC 方法。R. Xia 等人利用词性标注、词语关系等组合特征(Ensemble Features),针对朴素贝叶斯方法、最大熵模型以及支持向量机模型建立组合模型,实验效果优于传统的单一机器学习模型[18]。J. S. Deshmuke 等人将改进的最大熵模型与二分图聚类模型相结合,针对情感词分类取得了相对较高的准确度[19]。

基于深度学习的方法通常将源领域和目标领域的

词汇和文本以较低维度的向量(即词向量、句子向量或文档向量)来表示,并将其应用到深度学习算法(例如深度神经网络)之中。D. Tang 等人[20]较早地分析了深度学习方法在情感分析中的应用,发现其在情感分类、观点抽取和情感词典构建等方面优于传统方法。余传明等人[21]结合结构对等学习模型,将卷积神经网络模型应用到跨领域情感分析中,实验结果表明深度学习模型在较多指标上优于传统模型。J. Yu 等人[22]利用深度学习方法对句子进行建模,其实验结果表明,深度学习模型优于传统的结构对等学习模型。由于深度学习方法能够更好地揭示和获取不同领域文本信息的内在语义表示,从而使研究者脱离繁琐的传统特征工程,因而总体上优于传统的图模型方法和统计学习方法。

1.3 情感计算视角的跨领域文本情感分析

从情感计算的视角来看,跨领域文本情感分析主要包括词汇情感计算、篇章情感计算和句子情感计算等。跨领域词汇情感计算,是指在不同的领域情况下,判断词汇表达的情感是否为褒义(贬义)。例如,吴斐等人[23]将跨领域情感词汇计算分为三个步骤,即:基准词提取(提取具有非常明显的褒贬倾向的代表性词汇)、基准词消歧(判断基准词在不同领域的情感歧义)和目标词情感倾向判断(通过计算目标词和基准词的相关性强度来判断)。冯超等人[24]在深度学习的基础上,提出了一种基于词向量相似度的半监督情感词汇极性判断算法(Sentiment Orientation from Word Vector, SO-WV)。其实验结果表明,基于词向量的模型优于传统的点互信息(Pointwise Mutual Information, PMI)方法和标签传播(Label Propagation, LP)算法。

跨领域篇章情感计算是指在不同的领域情况下,判断文本篇章所表达的情感是否为褒义(贬义),其方法包括规则和统计两大类。在规则方面,K. Denecke 等人尝试利用领域通用的情感词汇词典(SentiWordNet)来进行跨领域篇章情感分析[25]。其实验结果表明,通过领域通用的情感词汇词典,结合一定的规则(例如,对篇章内的情感极性强度进行累加求和),在跨领域篇章情感分析上的效果弱于越过情感词汇词典而直接采用统计学习的方法;在统计方面,从简单的逻辑回归方法[16]和朴素贝叶斯方法[17]到复杂的支持向量分类方法和期望最大值方法[17],从单一学习模型[16-17]到组合学习模型[18]、从传统机器学习[19]到深度学习[22],跳过情感词汇词典而直接采用统计学习的方式成为跨领域篇章情感计算的主流。

跨领域句子情感计算是跨领域篇章情感计算的一种特殊情况(即将句子数量限制为1),其原理和方法与篇章层次类似,此处不再展开论述。综合在词汇、篇章和句子等层次跨领域文本情感分析方法,使用深度学习的方法总体上优于传统的基于词典的方法和基于规则的方法。

总体而言,跨领域相对于不跨领域,难点在于解决迁移学习问题;从非情感分析到情感分析,难点在于改进机器学习算法。由于深度学习模型能够更好地揭示和获取不同领域文本信息的内在语义表示,在解决迁移学习问题和改进机器学习算法等方面均取得了较大的优势。作为深度学习模型的一种重要代表,循环神经网络[26]近年来得到了越来越多的应用。典型的循环神经网络包括长段时记忆网络(Long Short Term Memory, LSTM)网络模型、双向长短时记忆网络(Bi-directional LSTM, Bi-LSTM)模型等。通过对文献进行搜索,目前尚未发现循环神经网络用于跨领域情感分析的实证研究。尽管如此,众多的实验表明,LSTM 能够有效地提升机器翻译[27]、语言建模[28]、多语言信息处理[29]以及图像的自动化标签(Automatic Image Captioning)[30]效果;Bi-LSTM 在中文分词[31]、句法分析[32]和词性标注[33]方面均能取得较好的效果。鉴于此,本文尝试将循环神经网络引入到跨领域文本情感分析的实证研究中,检验其在解决迁移学习问题和改进机器学习算法上的效果,以期为相关研究提供借鉴。

2 研究问题与方法

2.1 研究问题的形式化定义

本文的研究问题为:在给定源领域和目标领域大量无标注数据、以及源领域的少量标注数据的情况下,如何更好地利用迁移学习与深度学习理论来解决领域跨度环境下的概念漂移(特征漂移)问题,在篇章层次实现对目标领域的文本情感分析。具体而言,假定下标 s 表示源领域,下标 t 表示目标领域;D 表示领域集合;Train、Test、U 分别表示训练集、测试集和无标注文档,则上述问题可以描述为:给定源领域 D_s 中的有标注数据集 $Train_s$ 和无标注数据集 U_s,目标领域 D_t 中的测试集 $Test_t$ 和无标注文档 U_t,使用 $Train_s$、U_s、U_t 以及提出的模型实现对 $Test_t$ 的情感分类。

为简化描述,下文使用 $D_s \to D_t$ 表示源领域到目标领域的领域迁移学习。对于源领域和目标领域,为了便于与前人的研究结果进行对比,本文选择在跨领域研究中使用较多的三个领域,即:书籍、DVD 和音乐,

下文分别使用 Book、DVD 和 Music 来表示。

2.2 CD-DRNN 模型结构

本文在深度学习和迁移学习的理论基础上,提出 CD-DRNN 模型,其基本思路为:首先利用深度学习模型来进行跨领域表示学习,构建一个领域无关的特征空间,并将不同领域的特征表示投影到共同的空间中;通过堆叠式双向长短时记忆网络对源领域的标注数据进行监督学习,在每一层以正向和反向的顺序学习上一层的输出序列,在任意时刻 t,模型捕获正向和逆向的上下文信息,经过层层抽取的特征最终输入激活单元进行情感识别;将模型与参数共享到目标领域,对目标领域的无标注数据进行情感分析,并与正确结果进行对比。从方法和路线上来看,在知识迁移方面,采用模型迁移的方式,在源领域和目标领域共享堆叠式双向长短时记忆网络模型和参数;在机器学习方面,采用"深度表示学习 + 堆叠式双向长短时记忆网络模型"的组合方式,即利用无监督的预训练方式对源领域和目标领域的无标注数据进行表示学习,提取两个领域的共同语义特征,将共同语义特征加载到双向长短时记忆网络单元,多层逐步累计,最后作为激活层单元的输入;在情感计算方面,利用知识迁移与深度学习模型对跨领域文本在篇章层次上进行情感分析。

CD-DRNN 模型结构如图 1 所示,其架构包括输入层(Input Layer)、表示学习层(Representation Learning Layer)、双向长短时记忆网络层(Bi-directional Long Short Time Memory Layer)以及激活函数层(即输出层,Output Layer),以下分节展开论述。

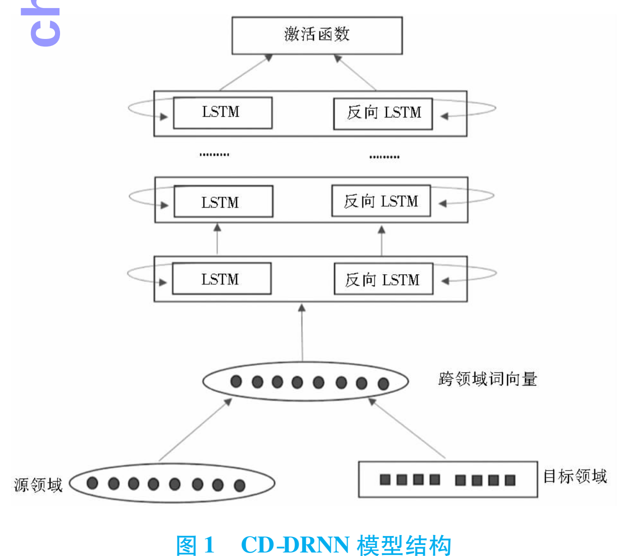


图 1 CD-DRNN 模型结构

2.2.1 输入层

在输入层,模型同时接受源领域和目标领域文档作为模型的输入。针对每个批处理大小(batch_size)的用户评论集合,取长度 s 作为当前输入文档集合的统一长度。如果长度文档小于 s,则以零填充。

2.2.2 表示学习层

在表示学习层,模型将源领域和目标领域文档中的词投射到共同的 d 维空间,即生成维度为 d 的词向量(Word Vector);同时,模型将每条评论视作一个文档,为其生成维度为 d 的文档向量(Doc Vector)。在表示学习过程中,模型利用词向量和文档向量(下文统一简称为 Word2Vec)来预测源领域和目标领域之中句子中出现的下一个单词。在本文实验中,使用平均作为组合向量的方法,利用上述建立的语料进行训练以获得相关的向量和参数。在模型启动阶段,文档向量和词向量被随机初始化,通过定义深度学习中的损失函数(即量化预测值与实际值之间的差距)和采用一定的优化方法(例如随即梯度下降方法,Stochastic Gradient Descending),最终获得文档向量和词向量作为上述预测任务的间接产物。假定语料库中包含 N 篇评论,词汇表中包含 M 个单词,则模型总共有 $N \times d + M \times d$ 参数。当 N 和 M 的值较大时,参数的数量可能也较大,参数更新在训练期间通常具有稀疏性。值得说明的是,词向量和文档向量同时从未标记的源领域和目标领域数据学习,能够获取两个领域的共同特征;特征的获取以通过上下文预测单词作为切入点,能够脱离繁重的人工标注过程。另外,在学习过程中,模型能够在小的语境中考虑到单词顺序,这点与 n-gram 模型方式相同,n-gram 模型保留了段落的大量信息,包括单词顺序。由于传统的 n-gram 模型往往需要创建一个非常高维的表示,而表示学习模型能够创建一个相对低维的表示,因此表示学习模型相比于传统的 n-gram 模型,具有更好的推广性能。

2.2.3 双向长短时记忆网络层

在针对源领域和目标领域所学习的共同特征的基础上,建立多层堆叠长短时记忆网络,每一层通过连接两个 LSTM(即正向 LSTM 和反向 LSTM)来完成建模。其中,正向 LSTM 与传统 LSTM 读取方式相同,反向 LSTM 则以逆序方式来读取输入序列。每个 LSTM 网络均由输入门、输出门、遗忘门和记忆单元构成。这些控制门和记忆单元的组合能够更好地表示和控制针对输入序列的长期、远期和近期的记忆,增强了模型处理远距离依赖的能力。在任意时刻 t,模型可以捕获到历史和未来的上下文信息。双向长短时记忆网络为多层,在第一层双向 LSTM 基础上,再叠加一层双向 LSTM,即:把第一层双向 LSTM 的输出作为第二层双向 LSTM 神经网络层相应

节点的输入;把第二层双向 LSTM 的输出作为第三层双向 LSTM 神经网络层相应节点的输入;以此类推。由于在相邻的层级之间,双向 LSTM 在结构上具有一一对应关系,因此从实验操作上能够较好地叠加。

2.2.4 输出层　输出层由激活单元构成,经过层层抽取的特征最终输入激活单元进行情感分类。通过激活单元得到一个 N 维的向量,其中,向量的第一个值为当前评论属于第一类的概率值;向量中的第二个值为当前评论属于第二类的概率值;后续以此类推。

在本文中,模型运行所需要设置的参数如表 1 所示:

表 1　CD-DRNN 模型参数设置

参数	取值
词向量维度	为了与对比算法中的模型保持一致,将词向量维度统一设置 100
文档表示维度	为了与对比算法中的模型保持一致,将文档向量维度统一设置 100
文档长度	考虑算法空间消耗和尽可能保留原始评论数据,采用评论长度的较大值 120
批处理大小	32
学习率	初始值为 0.1,使用 adam 优化算法在模型训练过程中,为参数分配不同的学习率
初始化方法	随机初始化方法
损失函数	使用交叉熵来定义损失函数
优化方法	adam
LSTM 网络激活函数	sigmoid
LSTM 网络输出维度	50
LSTM 网络层数	3
尾端激活函数	sigmoid
Epoch	训练阶段将模型训练 50 个 epoch
验证集	预训练时分别使用两个领域下无标注文档中的 10% 作为验证集

2.3　对比方法

本文将提出的模型与基线方法进行比较,选择了较具有代表性的 3 种单一机器学习模型(支持向量分类模型、逻辑回归模型和决策树模型)、3 种组合机器学习模型(随机森林模型、装袋模型和 Adaboost 模型)和 3 种循环神经网络模型(栈式长短时记忆模型、双向长短时记忆模型和卷积神经网络－长短时记忆网络组合模型(含串联和并联两种组合方式))。在特征抽取方面,选取了最具有代表性的传统词频逆文档频率方法,将其与深度学习中的词向量方法对比。在使用传统特征工程方法抽取源领域、目标领域的词汇过程中,为了使其与 CD-DRNN 及其他基线方法保持一致,使用主成分分析法将维度降为 100;在使用深度表示学习方法抽取源领域、目标领域的词汇过程中,为了使其与 CD-DRNN 及其他基线方法保持一致,统一将词向量与文档向量维度设置为 100。具体情况如下:

(1) 基于跨领域词频逆文档频度(Term Frequency Inverse Document Frequency, TFIDF)的机器学习方法。利用源领域和目标领域的训练集以及无标注评论集构建词典(词典大小为 93 000),并计算词汇的 TFIDF 权重;针对源领域训练集,利用主成分分析(Principal Component Analysis, PCA)[34]把特征维度从 93 000 维降为 100 维;把降维后的训练数据输入传统的机器学习模型,包括支持向量机分类(Support Vector Classification, SVC)[35]模型、逻辑回归(Logistic Regress, LR)[36]模型、决策树(Decision Tree, DT)[37]模型、随机森林(Random Forest, RF)[38]模型、装袋(Bagging)[39]模型以及 Adaboost[40]模型;利用主成分分析对目标领域的测试集进行降维,再利用训练好的机器学习模型对其进行分类。

(2) 基于跨领域词向量(Word Embedding, WE)[41]的机器学习方法。利用源领域和目标领域的训练集以及无标注评论集训练词向量(维度为 100);将源领域的评论转化为词向量作为训练数据输入传统的机器学习模型,包括 SVC、LR、DT、RF、Bagging 以及 Adaboost 等;针对目标领域的评论,利用训练好的机器学习模型输入词向量进行测试。

(3)栈式长短时记忆模型(Stacked Long Short Term Memory, Stacked-LSTM)[42]模型,即以多层的形式堆叠 LSTM 模型,底层 LSTM 的隐含状态输出作为高层 LSTM 的输入。将源领域的评论转化为词向量作为训练数据输入 Stacked-LSTM 模型;针对目标领域的评论,利用训练好的 Stacked-LSTM 模型输入词向量进行测试。

(4)双向长短时记忆模型(Bidirectional Long Short Term Memory, Bi-LSTM)[20]模型,即在 LSTM 模型的基础上添加反向(将输入逆序处理)学习过程。将源领域的评论转化为词向量作为训练数据输入 Bi-LSTM 模型;针对目标领域的评论,利用训练好的 Bi-LSTM 模型输入词向量进行测试。

(5)卷积神经网络与长短时记忆模型串联(Convolution Neural Network combined with Long Short Term Memory, CNN-LSTM)[43]模型,即在 LSTM 模型的前端添加卷积和池化过程,以提取更加准确的特征。将源领域的评论转化为词向量作为训练数据输入 CNN-LSTM 模型;针对目标领域的评论,利用训练好的 CNN-LSTM 模型输入词向量进行测试。

（6）卷积神经网络与长短时记忆并联（Convolution Neural Network merged with Long Short Term Memory，Merged-CNN-LSTM）模型，即在 LSTM 模型的侧端添加卷积和池化过程，将共同学习到的特征作为激活函数（Activation）的输入。将源领域的评论转化为词向量作为训练数据输入 Merged-CNN-LSTM 模型；针对目标领域的评论，利用训练好的 Merged-CNN-LSTM 模型输入词向量进行测试。

在本文中，基线方法运行所需要设置的参数如表 2 所示：

表 2 基线方法的参数设置

方法	参数
TFIDF + PCA + SVC（LR/DT/RF/Bagging/Adaboost）	在抽取源领域、目标领域的词汇过程中，形成词典大小为 93 000；利用主成分分析法降维，为了与其他基线方法保持一致，维度降为 100；SVC 方法中，惩罚参数 C 设置为 1，核函数为径向基核函数（Radial Basis Function，RBF）；LR 方法中，惩罚参数 C 设置为 1，惩罚方式设置为 l2；DT 方法中，特征选择标准为 gini，特征划分为在特征的所有划分点中找出最优的划分点；RF 方法中，最大的弱学习器的个数为 10；Bagging 方法中，基本估计器设置为 K 最近邻分类器（K-Nearest Neighbors，KNN）；Adaboost 方法中，最大的弱学习器个数为 100
WE + SVC（LR/DT/RF/Bagging/Adaboost）	利用源领域和目标领域的训练集以及无标注评论集训练跨领域词向量，为了与 CD-DRNN 模型进行对比，维度设置为 100；各种基线机器学习方法的参数设置与 TFIDF 情境下的设置相同
Stacked LSTM	词向量维度 100；文档向量维度 100；评论长度为 120；批处理大小为 32；优化方法为 adam；使用交叉熵（Cross Entropy）来定义损失函数；激活函数为 sigmoid；预训练时分别使用源领域和目标领域下无标注文档中的 10% 作为验证集
Bi-LSTM	词向量维度 100；文档向量维度 100；评论长度为 120；批处理大小为 32；优化方法为 adam；使用交叉熵来定义损失函数；激活函数为 sigmoid；训练阶段将模型训练 50 个 epoch；预训练时分别使用源领域和目标领域下无标注文档中的 10% 作为验证集
CNN-LSTM	卷积神经网络与长短时记忆网络采取串联的形式；卷积层采用一维卷积，卷积长度为 3，filter 数量为 1，激活函数为 Relu，步长为 1；池化层采用最大池化方法（Max Pooling），池化长度为 4；网络 Drop_out 比率为 0.2；分类层激活函数为 sigmoid；词向量维度 100；文档向量维度 100；评论长度为 120；批处理大小为 32；优化方法为 adam；使用交叉熵来定义损失函数；激活函数为 sigmoid；预训练时分别使用源领域和目标领域下无标注文档中的 10% 作为验证集
Merged-LSTM	卷积神经网络与长短时记忆网络采取并联的形式；卷积层采用一维卷积，卷积长度为 3，filter 数量为 1，激活函数为 Relu，步长为 1；池化层采用最大池化方法（Max Pooling），池化长度为 4；网络 Drop_out 比率为 0.2；分类层激活函数为 sigmoid；词向量维度 100；文档向量维度 100；评论长度为 120；批处理大小为 32；优化方法为 adam；使用交叉熵来定义损失函数；激活函数为 sigmoid；预训练时分别使用源领域和目标领域下无标注文档中的 10% 作为验证集

3 试验及分析

3.1 数据集

从亚马逊中文网站爬取书籍、音乐和 DVD 类目下的中文评论，每条用户评论对应由一星至五星表示的用户评分，该评分表达了用户对产品的满意程度。将三星以上的用户评论标注为正向评论，三星以下的评论标注为负向评论，去除三星评论，以此进行文本情感分析研究。每个领域中的评论划分为训练集、测试集和无标注文档。其中，书籍、音乐和 DVD 每个领域训练集包括 3 000 条评论（正负向情感类别各 1 500 条），测试集包括 1 000 条评论（正负向情感类别各 500 条），无标注集合共包括约130 000 条评论。

对三个领域的原始评论数据的长度进行了统计，结果如表 3 所示。其中，书籍领域评论的最大长度为 1 104，平均长度为 19；DVD 领域评论的最大长度为 741，平均长度为 19；音乐领域评论的最大长度为 759，平均长度为 18。从评论长度来看，书籍和 DVD 领域相同，音乐领域最短；从评论长度标准差来看，DVD 评论长度的变化较大，其次为书籍领域，音乐领域则变化最小。

表 3 各领域评论文本长度统计

领域	最大长度	最小长度	平均长度	长度标准差
Book	1104	1	19.0	33.00
DVD	741	1	19.0	34.77
Music	759	1	18.0	30.51

对三个领域的原始评论数据的长度占比进行了统计，结果如图 2 所示。可以看出，三个领域在各个区间上所占的比例呈现较强的一致性，具体而言：评论长度最多地分布在［0，10）和［10，30）两个区间，其次为［30，50）和［50，100），其他区间则相对较少。考虑算法空间消耗和尽可能保留原始评论数据，本文在后续各种模型（包括基线模型以及 CD-RNN 模型等）对评论长度取其较大值为 120。

3.2 实验结果

3.2.1 基于跨领域 TFIDF 的机器学习方法 本节采用受试者工作特征曲线（Receiver Operating characteristic Curve，ROC）以及曲线下面积（Area Under the Curve，AUC）来衡量各种方法的效果，各模型所得到的 ROC 曲线以及 AUC 面积如图 3 所示。由图 3 可以看出，在 Book→DVD、DVD→Book、Music→Book、Book→Music、Music→DVD 以及 DVD→Music 等六项跨领域情感分析实验中，LR 模型取得了最好的效果，其六项实

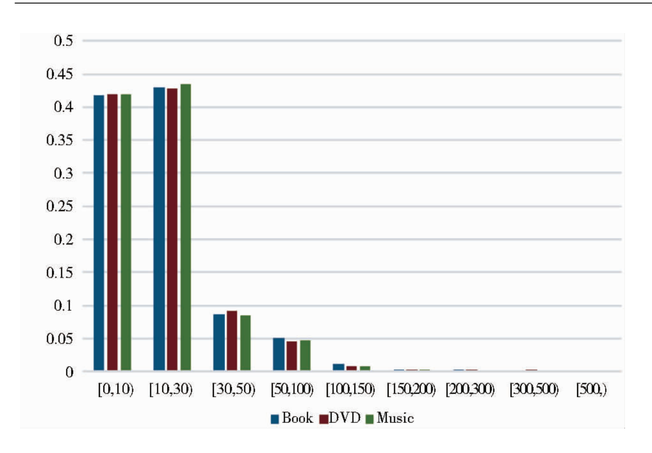


图 2　各领域文本评论长度占比统计

验所得到的 AUC 面积分别为 0.84、0.87、0.91、0.83、0.90 和 0.84(均值为 0.865);SVC 模型取得了仅次于 LR 模型的效果,其六项实验所得到的 AUC 面积分别为 0.82、0.86、0.91、0.81、0.88 和 0.83(均值为 0.852);居于第三位的是 Adaboost 模型,其六项实验所得到的 AUC 面积分别为 0.83、0.82、0.86、0.81、0.86 和 0.82(均值为 0.833);RF 模型和 Bagging 两种组合模型分别取得了排名第四位和第五位的效果,两种模型在六项实验所得的 AUC 面积均值分别为 0.812 和 0.757;DT 模型效果最弱,其六项实验所得到的 AUC 面积分别为 0.67、0.67、0.74、0.68、0.71 和 0.68(均值为 0.733)。

图 3　基于 TFIDF+PCA 的跨领域情感分析

从分组实验结果来看,各种算法在不同分组实验(即源领域和目标领域变化时)之间的适配效果呈现差异性。Music→Book 这一组实验在 SVC、LR、DT、RF、Bagging 以及 Adaboost 六种算法中均取得了最好的效果,排在其后的依次为 Music→DVD、DVD→Book、DVD→Music、Book→DVD 和 Book→Music。这表明源领域和目标领域的选择对跨领域的文本情感分析具有一定的影响;当源领域和目标领域发生变化时,模型的分类效果会产生相应改变。

3.2.2 基于跨领域词向量的机器学习方法 采用 ROC 和 AUC 来衡量各种模型的效果,各模型所得到的 ROC 曲线以及 AUC 面积如图 4 所示:

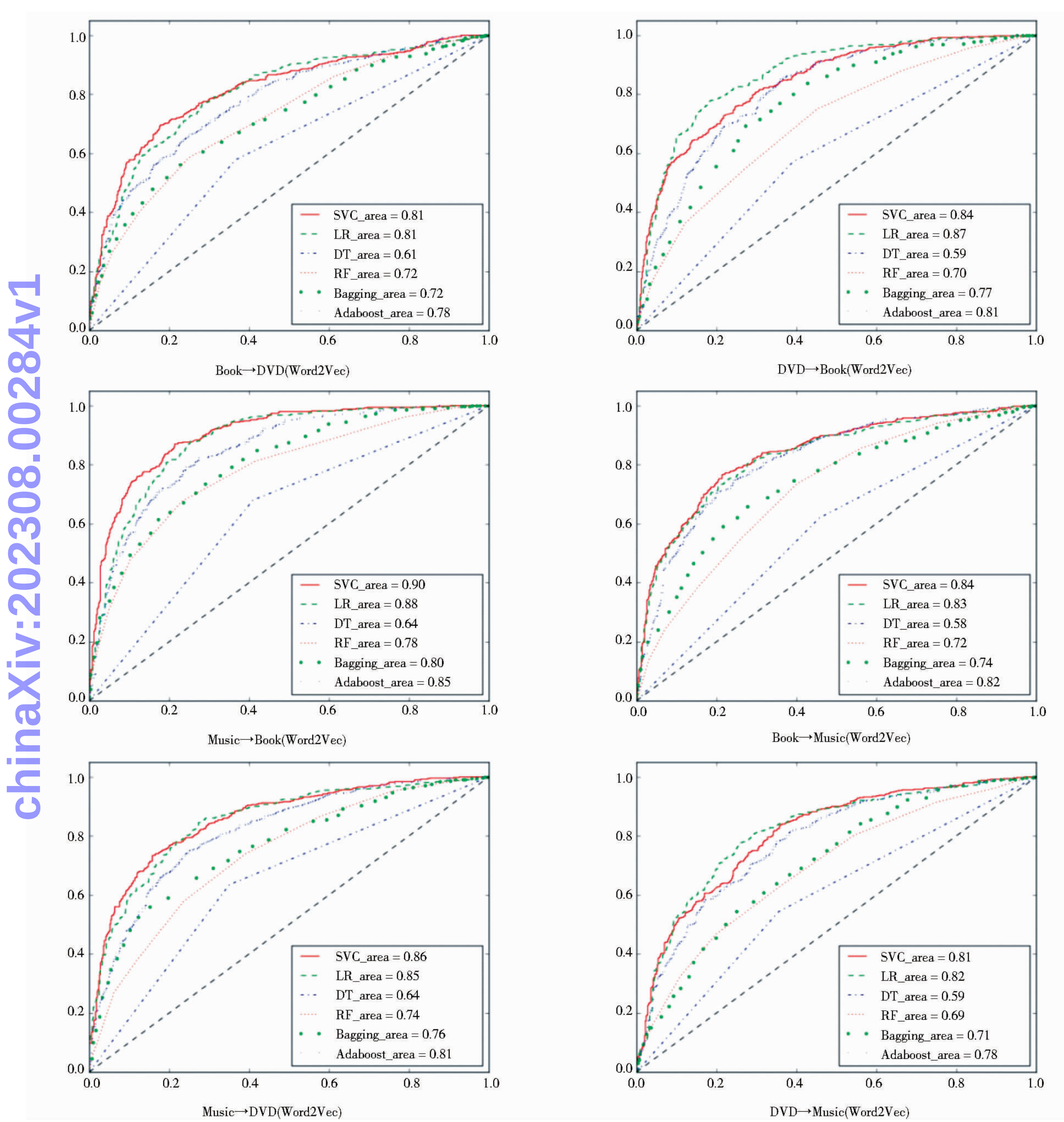


图 4 基于词向量的跨领域情感分析

从图 4 可以看出,在 Book→DVD、DVD→Book、Music→Book、Book→Music、Music→DVD 以及 DVD→Music 等六项跨领域情感分析实验中,SVC 模型和 LR 模型取得了最好的效果,二者在六项实验所得到的 AUC 面积均值都为 0.843; Adaboost 模型取得了仅次于 SVC 和 LR 模型的效果,其六项实验所得到的 AUC 面积分别为 0.78、0.81、0.85、0.82、0.81 和 0.78(均值为 0.808);居于第四位的是 Bagging 模型,其六项实验所得到的 AUC 面积分别为 0.72、0.77、0.80、0.74、0.76 和 0.71(均值为 0.750);居于第五位的是 RF 模型,其

六项实验所得到的 AUC 面积分别为 0.72、0.70、0.78、0.72、0.74 和 0.69(均值为 0.725);DT 模型效果最弱,其六项实验所得到的 AUC 面积分别为 0.61、0.59、0.64、0.58、0.64 和 0.59(均值为 0.608)。

对比图 3 和图 4 的结果,可以看出,相对于基于 TFIDF 的跨领域机器学习模型,基于词向量的方法未能有效提升跨领域机器学习的效果。在 Music→Book、Book→DVD 和 DVD→Music 等分组测试中,基于词向量在 SVC、LR、DT、RF、Bagging 以及 Adaboost 六种算法上的效果还有所下降(相较于基于 TFIDF 的方法)。从图 3 和图 4 的对比实验结果可以看出,仅仅通过跨领域的词向量学习,并不能十分有效地获取两个领域的共同特征。在传统的跨领域词向量学习的基础上,有必要加入引入更多的特征抽取机制(例如多层双向长短时记忆网络等),来提升跨领域知识迁移与特征映射的效果。

3.2.3 基于循环神经网络模型的方法 由于循环神经网络模型通常以批处理(Batch-wise)形式来进行测试,AUC 和 ROC 的值通常取决于测试集合中各分块(Batch)的随机划分。为了使测试结果具有更好的可再现性(Reproducibility),在循环神经网络部分,采用准确度(Accuracy, Acc)来衡量实验效果。

表 4 显示了各种循环神经网络模型(包括 Stacked-LSTM、Bi-LSTM、CNN-LSTM、Merged-CNN-LSTM 以及 CD-DRNN)、跨领域 TFIDF 机器学习模型(包括 TFIDF + PCA + SVC、TFIDF + PCA + LR、TFIDF + PCA + DT、TFIDF + PCA + RF、TFIDF + PCA + Bagging 和 TFIDF + PCA + Adaboost)和跨领域词向量机器学习模型(包括 WE + SVC、WE + LR、WE + DT、WE + RF、WE + Bagging 和 WE + Adaboost)在 Book→DVD、DVD→Book、Music→Book、Book→Music、Music→DVD 以及 DVD→Music 等六项跨领域情感分析实验中的结果。在表 4 中,黑色加粗字体部分为每一项实验通过上述各种方法所取得的最大值,最右一列表示各种算法在六项实验中所取得的平均值。

由表 4 可以看出,在 DVD→Book、Music→Book、Music→DVD 以及 DVD→Music 四项试验中,CD-DRNN 取得最高的 Acc 值,分别为 81.50%、86.10%、85.30% 和 78.20%;在 Book→DVD 以及 Book→Music 两项实验中,Merged-CNN-LSTM 和 Stacked-LSTM 取得了最高的 Acc 值,分别为 82.50% 和 78.50%。从六项实验取得的平均值来看,各种循环神经网络模型按照效果从高到低排列依次为 CD-DRNN(81.70%)、Merged-CNN-LSTM(80.90%)、Bi-LSTM(80.50%)、Stacked-LSTM(79.90%)、CNN-LSTM(74.70%),这表明本文所提出的模型能够更好地进行跨领域的文本情感分析,从而验证了该模型有效性。

表 4 跨领域情感分析的比较研究结果

算法\领域变化	Book→DVD	DVD→Book	Music→Book	Book→Music	Music→DVD	DVD→Music	Average
TFIDF + PCA + SVC	0.517	0.513	0.556	0.517	0.563	0.521	0.531
TFIDF + PCA + LR	0.763	0.753	0.841	0.757	0.809	0.706	0.772
TFIDF + PCA + DT	0.674	0.675	0.736	0.681	0.707	0.683	0.693
TFIDF + PCA + RF	0.728	0.752	0.807	0.704	0.750	0.692	0.739
TFIDF + PCA + Bagging	0.684	0.672	0.715	0.657	0.703	0.663	0.682
TFIDF + PCA + Adaboost	0.747	0.733	0.799	0.729	0.774	0.721	0.751
WE + SVC	0.680	0.748	0.785	0.724	0.758	0.708	0.734
WE + LR	0.743	0.786	0.808	0.762	0.778	0.733	0.768
WE + DT	0.606	0.590	0.635	0.581	0.642	0.594	0.608
WE + RF	0.668	0.638	0.722	0.644	0.671	0.634	0.663
WE + Bagging	0.605	0.694	0.673	0.603	0.625	0.642	0.640
WE + Adaboost	0.700	0.732	0.772	0.744	0.748	0.698	0.732
Stacked-LSTM	0.802	0.787	0.855	0.785	0.834	0.729	0.799
Bi-LSTM	0.823	0.797	0.827	0.759	0.840	0.781	0.805
CNN-LSTM	0.744	0.736	0.819	0.730	0.740	0.711	0.747
Merged-CNN-LSTM	0.825	0.810	0.861	0.785	0.828	0.744	0.809
CD-DRNN	0.818	0.815	0.861	0.772	0.853	0.782	0.817

对比循环神经网络模型、跨领域 TFIDF 机器学习模型和跨领域词向量机器学习模型,可以发现,循环神经网络模型的效果普遍优于跨领域 TFIDF 机器学习模型和跨领域词向量机器学习模型。从六项实验取得的平均值来看,效果最好的是本文所提出的深度循环神经网络模型(CD-CRNN, 81.70%),它比最好的跨领域 TFIDF 机器学习模型(TFIDF + PCA + LR,77.20%)高 4.50%,比最好的跨领域词向量机器学习模型(WE + LR,76.80%)高 4.90%。

3.3 讨论

对比上述三组实验可以看出,本文所提出的基于深度循环神经网络的方法在跨领域文本情感分析的效果优于各种基线方法。相对于传统的 TFIDF 方法,该模型在六组实验中的均值最高提升约 28.6%(对比表 4 第一行与最后一行);相对于基于词向量的方法,该模型在六组实验中的均值最高提升约 20.9%(对比表 4 第九行与最后一行);相对于目前所用的循环神经网络,该模型在六组试验中的均值最高提升约 7%(对比表 4 倒数第三行与最后一行)。与基于词向量的方法未能有效提升跨领域机器学习的效果(对比图 3 和图 4)不同,CD-DRNN 模型与基线的对比表明该模型能够有效提升跨领域文本情感识别的有效性。这说明通过堆叠多层 LSTM 单元,让每一层 LSTM 单元分别以正向和反向的顺序学习上一层的输出序列,能够有效地获取情感分类在多个领域的有效特征,从而更有效地解决了不同领域之间的概念漂移问题,大大提升了模型的分类效果。

从分组实验结果来看,各种算法在不同分组实验(即源领域和目标领域变化时)之间的适配效果呈现差异性。例如,Music→Book 这一组实验在几乎所有算法中均取得了最好的效果(从 Acc 来看),排在其后的依次为 Music→DVD、Book→DVD、DVD→Book、DVD→Music 和 Book→Music。可以看出,源领域与目标领域发生变化时,各种模型的识别效果都会有较大的变化,这表明源领域和目标领域的选择对跨领域的文本情感分析具有一定的影响。通常认为,领域之间的相似程度越高,则模型的领域适配效果会越好[9]。尽管本文对领域的相似程度与模型适配效果的正相关性未作出检验,但通过对比 Book→Music 和 Music→Book 两组实验在不同的算法之中的效果可以看出,两组实验在各组算法的效果均存在较大的差异。这表明,在度量源领域和目标领域的相似性时,有必要考虑领域之间所存在的非对称性。

从数据使用角度来看,本文所使用的标注数据数量仅为每个领域 4 000 条(即正向数据和负向数据各 2 000 条),远少于无标注数据的数量(130 000 条)。在标注数量较小的情况下,模型所取得的文本情感分类效果仍然好于基线方法。这表明,通过 CD-DRNN 模型从无标注数据中有效地筛选特征,能够大大降低源领域和目标领域数据标注相关的工作量。值得说明的是,CD-DRNN 模型同时从无标记的源领域和目标领域文档中学习,特征的获取以通过上下文预测单词作为切入点,从而使得模型应用能够脱离繁重的人工标注过程。

从模型的内在机理来看,本文涉及知识迁移、机器学习与情感计算等视角。在知识迁移方面,CD-DRNN 采用模型迁移的方式,在源领域和目标领域共享堆叠式双向长短时记忆网络模型和参数。模型迁移方法减少了传统特征迁移方法对枢纽词汇(领域无关词汇)和样本迁移方法对源领域标注数据的依赖,使模型显得更加简洁。在机器学习方面,CD-DRNN 模型采用"深度表示学习 + 堆叠式双向长短时记忆网络模型"的组合方式,其中,深度表示学习模型能够更好地揭示和获取不同领域文本信息的内在语义表示;相对于传统的循环神经网络模型,记忆单元的加入能够更好地表示和控制针对输入序列的长期、远期和近期的记忆,增强了处理远距离依赖的能力,有效地解决传统循环神经网络在处理长度较大的序列数据时所面临的梯度消失和梯度爆炸问题;模型同时考虑正向读取方式和反向读取方式,能够有效地解决反向依赖问题;模型采用多层堆叠,能够在多个层次上提取与领域相关的情感极性特征,综合使得模型在跨领域情感分析的准确度上优于基线机器学习方法。在情感计算方面,利用知识迁移与机器学习模型对跨领域文本在篇章层次上进行情感分析,摆脱了对于建立跨领域情感词典的依赖,因而更易于推广到其他应用。

本文与同类研究,如文[21]相比,两者既存在一定关联,也存在本质区别。首先,从研究方法来看,文[21]采用深度卷积神经网络来进行跨领域情感分析,而本文则采用深度循环神经网络。两种网络在原理上存在本质差异,前者通过卷积、池化等操作来逐层提取特征,后者通过多层堆叠式双向长短时记忆网络来提取特征,在深度学习领域隶属于不同的研究范畴。其次,从实验设计来看,由于本文所提出模型的理论基础不同于文[21],因此采用不同的基线方法和评价指标。文[21]提出基于卷积神经网络的模型,于是与基于结构对等模型的分布对应方法这种基线方法在 F1 指标上进行比对;本文提出基于循环神经网络的模型,与九种基线方法(即三种单一机器学习模型、三种组合机器学习模型和三种循环神经网络模型)在 AUC 指标和 ACC 指标上进行比对。再次,从模型内在机理来看,本文采用了不同的知识迁移方法。文[21]依赖于特征迁移(领域无关词汇),使用源领域和目标领域的枢纽词汇来表示文档中的词项;本文采用模型迁移的方式,在源领域和目标领域共享堆叠式双向长短时记忆网络模型和参

数,减少了枢纽词汇和样本迁移方法对源领域标注数据的依赖,使模型显得更加简洁。最后,从实验结果来看,与文[21]相比,本文提出的模型在与多数基线方法的对比实验中保持了显著的优势,实验结果更加有力地证明了深度循环神经网络模型在跨领域文本情感分析中的有效性。

4 结语

本文选择亚马逊在书籍、DVD 和音乐类目下的中文评论作为实验数据,情感分析作为研究任务,提出了一种面向跨领域情感分析的深度循环神经网络模型。该模型在跨领域环境下的平均分类准确度达到了 81.70%,优于传统的栈式长短时记忆网络、双向长短时记忆网络模型模型、卷积神经网络长短时记忆网络串联模型以及卷积神经网络长短时记忆网络并联模型,实验结果证明了模型的有效性。

本文的局限性在于语料库的选择主要局限于中文。在后续研究中,将在多语言语料库中验证模型,并在领域跨度更大的环境中(例如电子产品与餐饮类评论)中进一步检验模型的效果;同时,将利用更多的深度学习模型(例如生成对抗网络模型等)来进一步丰富跨领域文本情感分析的相关研究。

参考文献:

[1] WEISS K, KHOSHGOFTAAR T M, WANG D D. A survey of transfer learning[J]. Journal of big data, 2016, 3(1):1 - 40.
[2] TAHMORESNEZHAD J, HASHEMI S. Visual domain adaptation via transfer feature learning[J]. Knowledge and information systems, 2017, 50(2):1 - 21.
[3] BIFET A, PECHENIZKIY M, BOUCHACHIA A. A survey on concept drift adaptation[J]. ACM computing surveys, 2014, 46(4):1 - 44.
[4] BARDDAL J P, GOMES H M, ENEMBRECK F, et al. A survey on feature drift adaptation: definition, benchmark, challenges and future directions[J]. Journal of systems and software, 2016,127:278 - 294.
[5] VAPNIK V, IZMAILOV R. Knowledge transfer in SVM and neural networks[J]. Annals of mathematics and artificial Intelligence, 2017,80(1 - 2):3 - 19.
[6] 徐军, 丁宇新, 王晓龙. 使用机器学习方法进行新闻的情感自动分类[J]. 中文信息学报, 2007, 21(6):95 - 100.
[7] 吴应良, 黄媛, 王选飞. 在线中文用户评论研究综述:基于情感计算的视角[J]. 情报科学, 2017(6):159 - 163.
[8] BLITZER J, MCDONALD R, PEREIRA F. Domain adaptation with structural correspondence learning[C]// Conference on empirical methods in natural language processing. Stroudsburg, PA, USA:Association for Computational Linguistics, 2006:120 - 128.
[9] PAN S J, NI X, SUN J T, et al. Cross-domain sentiment classification via spectral feature alignment[C]// International conference on World Wide Web, WWW 2010, Raleigh, North Carolina, U. S. A., April. DBLP, 2010:751 - 760.
[10] DAI W, YANG Q, XUE G R, et al. Boosting for transfer learning [C]// International conference on machine learning. New York, NY, USA:ACM, 2007:193 - 200.
[11] 马凤闸, 吴江宁, 杨光飞. 基于双重选择策略的跨领域情感倾向性分析[J]. 情报学报, 2012, 31(11):1202 - 1209.
[12] GLOROT X, BORDES A, BENGIO Y. Domain adaptation for large - Scale sentiment classification: a deep learning approach [C]// International conference on machine learning. Madison, WI, USA: Omnipress,2011:1 - 9.
[13] SUN M, TAN Q, DING R, et al. Cross-domain sentiment classification using deep learning approach[C]// International conference on cloud computing and intelligence systems. New York, NY, USA:IEEE, 2015:60 - 64.
[14] 吕韶华, 杨亮, 林鸿飞. 基于 SimRank 的跨领域情感倾向性分析算法研究[J]. 中文信息学报, 2012, 26(6):38 - 44.
[15] 魏现辉, 张绍武, 杨亮, 等. 基于加权 SimRank 的跨领域文本情感倾向性分析[J]. 模式识别与人工智能, 2013, 26(11):1004 - 1009.
[16] 张志武. 跨领域迁移学习产品评论情感分析[J]. 现代图书情报技术, 2013(6):49 - 54.
[17] 黄瑞阳, 康世泽. 一种改进 EM 算法的跨领域情感分类方法[J]. 计算机应用研究, 2017, 34(9):2696 - 2699.
[18] XIA R, ZONG C, HU X, et al. Feature ensemble plus sample selection: domain adaptation for sentiment classification[J]. IEEE intelligent systems, 2013, 28(3):10 - 18.
[19] DESHMUKH J S, TRIPATHY A K. Entropy based classifier for cross-domain opinion mining[J]. Applied computing and informatics, 2017,14(1):55 - 64.
[20] TANG D, QIN B, LIU T. Deep learning for sentiment analysis: successful approaches and future challenges[J]. Wiley interdisciplinary reviews data mining and knowledge discovery, 2015, 5 (6):292 - 303.
[21] 余传明, 冯博琳, 安璐. 基于深度表示学习的跨领域情感分析[J]. 数据分析与知识发现, 2017(7):73 - 81.
[22] YU J, JIANG J. Learning sentence embeddings with auxiliary tasks for cross-domain sentiment classification[C]// Conference on empirical methods in natural language processing. Stroudsburg, PA, USA:ACL,2016:236 - 246.
[23] 吴斐, 张玉红, 胡学钢. 面向评论信息的跨领域词汇情感倾向判别方法[J]. 计算机科学, 2015, 42(6):220 - 222.
[24] 冯超, 梁循, 李亚平, 等. 基于词向量的跨领域中文情感词典构建方法[J]. 数据采集与处理, 2017, 32(3):579 - 587.
[25] DENECKE K. Are SentiWordNet scores suited for multi-domain sentiment classification? [C]// International conference on digital information management. New York, NY, USA:IEEE, 2009:1 - 6.

[26] GRAVES A. Supervised sequence labelling with recurrent neural networks[M]. Berlin, Heidelberg: Springer, 2012.

[27] SUTSKEVER I, VINYALS O, Le Q V. Sequence to sequence learning with neural networks [EB/OL]. [2017 - 06 - 30]. https://arxiv.org/pdf/1409.3215.

[28] JOZEFOWICZ R, VINYALS O, SCHUSTER M, et al. Exploring the limits of language modeling [EB/OL]. [2017 - 06 - 30]. https://arxiv.org/pdf/1602.02410.

[29] DAN G, BRUNK C, VINYALS O, et al. Multilingual language processing from bytes [EB/OL]. [2017 - 06 - 30]. https://arxiv.org/pdf/1512.00103.

[30] VINYALS O, TOSHEV A, BENGIO S, et al. Show and tell: a neural image caption generator [C] // IEEE Conference on Computer Vision and Pattern Recognition. New York, NY, USA: IEEE, 2015: 3156 - 3164.

[31] YAO Y, HUANG Z. Bi-directional LSTM recurrent neural network for Chinese word segmentation [C] // International Conference on Neural Information Processing. Cham, Switzerland: Springer, 2016: 345 - 353.

[32] CROSS J, HUANG L. Incremental parsing with minimal features using Bi-Directional LSTM[C] // Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA, USA: ACL, 2016: 32 - 37.

[33] PLANK B, SØGAARD A, GOLDBERG Y. Multilingual part-of-speech tagging with bidirectional long short-term memory models and auxiliary loss[C] // Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA, USA: ACL, 2016: 1 - 7.

[34] ABDI H, WILLIAMS L J. Principal component analysis[J]. Wiley interdisciplinary reviews computational statistics, 2010, 2(4): 433 - 459.

[35] ULLRICH C. Support vector classification[M]// Forecasting and hedging in the foreign exchange markets. Berlin Heidelberg: Springer, 2009: 345 - 356.

[36] MENARD S. Applied logistic regression analysis[J]. Technometrics, 2002, 38(2): 184 - 186.

[37] LANDGREBE D. A survey of decision tree classifier methodology [J]. IEEE transactions on systems man and cybernetics, 2002, 21(3): 660 - 674.

[38] BREIMAN L. Random Forest[J]. Machine learning, 2001, 45(1): 5 - 32.

[39] JORDAN S J, VIVIAN A, WOHAR M E. Forecasting market returns: bagging or combining? [J]. International journal of forecasting, 2017, 33(1): 102 - 120.

[40] WANG J, GAO L, ZHANG H, et al. Adaboost with SVM-based classifier for the classification of brain motor imagery tasks[C]// International conference on universal access in human-computer interaction: users diversity. Berlin Heidelberg: Springer - Verlag, 2011: 629 - 634.

[41] LAI S, LIU K, HE S, et al. How to generate a good word embedding[J]. IEEE intelligent systems, 2016, 31(6): 5 - 14.

[42] DYER C, BALLESTEROS M, WANG L, et al. Transition-based dependency parsing with stack long short-term memory[J]. Computer science, 2015, 37(2): 321 - 332.

[43] WANG J, YU L C, LAI K R, et al. Dimensional sentiment analysis using a regional CNN-LSTM model[C]// Meeting of the Association for Computational Linguistics. Stroudsburg, PA, USA: ACL, 2016: 225 - 230.

A Cross-domain Text Sentiment Analysis Based on Deep Recurrent Neural Network

Yu Chuanming
School of Information and Safety Engineering, Zhongnan University of Economics and Law, Wuhan 430073

Abstract: [**Purpose/significance**] In order to solve the problem of classification model in target domain that caused by the lack of data, this study firstly trains the model of source domain that includes rich labeling/tagging data, and then, projects source and target domain documents into the same feature space. [**Method/process**] The reviews of three product categories, i.e. books, DVD and music, from Amazon, which are written in Chinese, are taken as the experimental data, and the cross-domain text sentiment analysis is considered as the research task. A novel model, i.e. the Cross Domain Deep Recurrent Neural Network (CD-DRNN), is proposed to achieve knowledge transfer among domains. The average accuracy value of CD-DRNN achieves 81.70%, which excels the values of Stacked Long Short Term Memory (79.90%), Bidirectional Long Short Term Memory (80.50%), Convolution Neural Network with Long Short Term Memory (74.70%) and Merged Convolution Neural Network with Long Short Term Memory (80.90%). [**Result/conclusion**] Knowledge transfer in source domain and target domain could effectively solve the difficulties of achieving good classification performances on small data sets. The proposed method can be leveraged to effectively select features from unlabeled data, thereby greatly reducing the workload related to data annotation in the target domain.

Keywords: cross domain transfer learning deep recurrent neural network sentiment analysis